

IN THE CLAIMS

Please amend the following claims:

- Dr
Sub
8/14
1. (Amended) A method for document classification comprising:
analyzing textual content and graphical content of a previously unclassified electronic document to determine a textual profile and a graphical profile of the electronic document;
generating a classification of the document based on the textual profile and the graphical profile;
storing the electronic document in a first directory structure based on the classification of the document and a document classification profile associated with the first directory structure; and
wherein the first directory structure mirrors a second directory structure, and the document classification profile is defined according to a classification approach of a user with respect to the second directory structure.
 2. (Amended) The method defined in Claim 1 where the first directory structure comprises a hierarchy of documents mirroring in a similar fashion an organization in the second directory structure representing a pre-existing memory storing documents.
 3. The method defined in Claim 2 when the pre-existing memory comprises a hard disk.

4. The method of claim 1, wherein analyzing textual content of an electronic document comprises:

determining characteristic words of the document;
determining a frequency for each characteristic word; and
building a frequency table based on the frequency associated with each characteristic word.

5. The method of claim 1, wherein analyzing graphical content of an electronic document comprises:

determining a point set corresponding to the electronic document, wherein points of the point set correspond to end points of lines;
determining a density of points within the point set;
generating a document profile based, at least in part, on the density of points within the point set.

6. The method of claim 1, wherein generating a classification of the document based on the textual and graphical properties comprises combining results from the textual and graphical analysis using a Borda Count.

7. (Amended) The method defined in Claim 1 further comprising building the first directory structure by building hierarchy of documents based on a user's hard drive.

8. (Amended) The method defined in Claim 1 further comprising building the first directory structure by extracting graphical and text features from documents in a directory-based memory to obtain a document classification profile of each subdirectory in the directory-based memory.

9. (Amended) A software product including a machine-readable medium having stored thereon sequences of instructions, which, when executed by a processor, cause the processor to:

analyze textual content and graphical content of a previously unclassified electronic document to determine a textual profile and a graphical profile of the electronic document;

generate a classification of the document based on the textual profile and the graphical profile;

store the electronic document in a first directory structure based on the classification of the document and a document classification profile associated with the first directory structure; and

wherein the first directory structure mirrors a second directory structure, and the document classification profile is defined according to a classification approach of a user with respect to the second directory structure.

10. The machine-readable medium of claim 9, wherein the sequences of instructions that cause the processor to analyze textual content of an electronic document further comprise sequences of instructions that cause the processor to:

determine characteristic words of the document;

determine a frequency for each characteristic word; and
build a frequency table based on the frequency associated with each characteristic word.

11. The machine-readable medium of claim 9, wherein the sequences of instructions that cause the processor to analyze graphical content of an electronic document further comprise sequences of instructions that cause the processor to:

determine a point set corresponding to the electronic document, wherein points of the point set correspond to end points of lines;

determine a density of points within the point set;

generate a document profile based, at least in part, on the density of points within the point set.

12. The machine-readable medium of claim 9, wherein the sequences of instructions that cause the processor to generate a classification of the document based on the textual and graphical properties further comprises sequences of instructions that cause the processor to combine results from the textual and graphical analysis using a Borda Count.

13. (Amended) A method for document classification comprising:
analyzing documents in a pre-existing document directory structure to determine a document classification profile of the pre-existing document directory structure, the document classification profile of the pre-existing document directory structure reflecting

a classification approach of a user with respect to the pre-existing document directory structure;

generating a mirror directory structure based on the pre-existing document directory structure;

receiving a previously unclassified electronic document;

analyzing textual content and graphical content of the electronic document to determine a textual profile and a graphical profile of the electronic document; and

placing the electronic document in the mirror directory structure based on the document classification profile of the pre-existing document directory structure, the textual profile of the document, and the graphical profile of the document to resemble the classification approach of the user.

14. The method of claim 13, wherein analyzing the pre-existing document directory structure further comprises:

recursively descending the pre-existing document directory structure;

generating a list of directories in the pre-existing document directory structure;

examining files in directories of the pre-existing document directory structure to determine content; and

examining the content of the files to determine document classification profile of the directories in the pre-existing document directory structure.

15. The method of claim 13 wherein the pre-existing document directory structure is a hard disk directory structure.

16. The method of claim 13 wherein generating a mirror directory structure based on the pre-existing document directory structure comprises generating a document directory structure having a set of directories and relationships equivalent to the pre-existing document directory structure.

17. The method of claim 13 wherein placing the electronic document in the mirror directory structure comprises:

determining a primary directory in the pre-existing document directory structure in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

storing the document in a primary corresponding directory in the mirror directory structure that corresponds to the primary directory in the pre-existing document directory structure.

18. The method of claim 17 further comprising:

determining a secondary directory in the pre-existing document directory in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

storing the document in a corresponding secondary directory in the mirror directory structure that corresponds to the secondary directory in the pre-existing document directory structure.

19. (Amended) A computer-readable medium having stored thereon sequences of instructions which, when executed by a processor, cause the processor to:

analyze a pre-existing document directory structure to determine a document classification profile of the pre-existing document directory structure, the document classification profile of the pre-existing document directory structure reflecting a classification approach of a user with respect to the pre-existing document directory structure;

generate a mirror directory structure based on the pre-existing document directory structure;

receive a previously unclassified electronic document;

analyze textual content and graphical content of the electronic document to determine a textual profile and a graphical profile of entire electronic document; and

place the electronic document in the mirror directory structure based on the document classification profile of the pre-existing document directory structure, the textual profile of the document, and the graphical profile of the document to resemble the classification approach of the user.

20. The computer-readable medium of claim 19, wherein the sequences of instructions that cause the processor to analyze a pre-existing document directory structure to determine an organization of the pre-existing document directory structure further comprise sequences of instructions that cause the processor to:

recursively descending the pre-existing document directory structure;

generating a list of directories in the pre-existing document directory structure;

examining files in directories of the pre-existing document directory structure to determine content; and

examining the content of the files to determine the organization of the directories in the pre-existing document directory structure.

21. The computer-readable medium of claim 19, wherein the sequences of instructions that cause the processor to generate a mirror directory structure further comprise sequences of instructions that cause the processor to generate a document directory structure having a set of directories and relationships equivalent to the pre-existing document directory structure.

22. The computer-readable medium of claim 19, wherein the sequences of instructions that cause the processor to place a document in the mirror directory structure further comprise sequences of instructions that cause the processor to:

determine a primary directory in the pre-existing document directory structure in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

store the document in a primary corresponding directory in the mirror directory structure that corresponds to the primary directory in the pre-existing document directory structure.

23. The computer-readable medium of claim 22 further comprising sequences of instructions that cause the processor to:

determine a secondary directory in the pre-existing document directory in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

store the document in a corresponding secondary directory in the mirror directory structure that corresponds to the secondary directory in the pre-existing document directory structure.

24. (Amended) An apparatus comprising:

means for analyzing a pre-existing document directory structure to determine document classification profile of the pre-existing document directory structure, the document classification profile of the pre-existing document directory structure reflecting a classification approach of a user with respect to the pre-existing document directory structure;

means for generating a mirror directory structure based on the pre-existing document directory structure;

means for receiving a previously unclassified electronic document;

means for analyzing textual content and graphical content of the electronic document to determine a textual profile and a graphical profile of the electronic document; and

means for placing the electronic document in the mirror directory structure based on the document classification profile of the pre-existing document directory structure, the textual profile of the document, and the graphical profile of the document to resemble the classification approach of the user.

25. The apparatus of claim 24, wherein means for analyzing the pre-existing document directory structure further comprises:

means for recursively descending the pre-existing document directory structure;

means for generating a list of directories in the pre-existing document directory structure;

means for examining files in directories of the pre-existing document directory structure to determine content; and

means for examining the content of the files to determine document classification profile of the directories in the pre-existing document directory structure.

26. The apparatus of claim 24, wherein means for generating a mirror directory structure comprises means for generating a document directory structure having a set of directories and relationships equivalent to the pre-existing document directory structure.

27. The apparatus of claim 24, wherein means for placing a document in the mirror directory structure comprises:

means for determining a primary directory in the pre-existing document directory structure in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

means for storing the document in a primary corresponding directory in the mirror directory structure that corresponds to the primary directory in the pre-existing document directory structure.

28. The apparatus of claim 27 further comprising:

means for determining a secondary directory in the pre-existing document directory in which the document should be placed based on the document classification profile of the pre-existing document directory structure; and

means for storing the document in a corresponding secondary directory in the mirror directory structure that corresponds to the secondary directory in the pre-existing document directory structure.

29. (Amended) A document processing system comprising:

a document scanning device;

a document storage device coupled to the document scanning device, wherein the document storage device is organized as a document directory structure having multiple directories, and further wherein the document storage device has a mirror directory structure having multiple directories organized based on the document directory structure; and

a processor coupled to the document scanning device and to the document storage device, wherein the processor is to analyze content of a document scanned by the document scanning device, to determine a directory in the mirror directory structure, in which the document will be placed, based on the analysis of document content and a document classification profile of the document directory structure, the document classification profile being defined according to a classification approach of a user with respect to the document directory structure, and to store the document in the directory in the mirror directory structure.

30. The document processing system of claim 29 wherein the processor is operable to determine a secondary directory in the document directory structure in which the document should be placed and to store the document in a corresponding secondary directory in the mirror directory structure.

31. The document processing system of claim 29 wherein the processor analyzes files stored in the document directory structure to determine content and generates the document classification profile of the document directory structure based on the analysis.

32. The document processing system of claim 29 wherein the document is analyzed based on image and textual content.